A FORTIORI BAYESIAN INFERENCE IN PSYCHOLOGICAL RESEARCH

Milton L. Lavin

September 1970 #478-70

A FORTIORI BAYESIAN INFERENCE
IN PSYCHOLOGICAL RESEARCH

Milton L. Lavin

September 1970 #478-70

Abstract

Bayes Law, or the law of conditional probability, provides a natural inference framework for one who views hypothesis testing as parameter estimation. Heretofore a major difficulty in applying Bayesian ideas to psychological contexts has been the specification of an objective or public prior. This paper proposes a rule for selecting a prior hypothesis which is both unambiguous and hostile to the research hypothesis: choose the prior so that 1) its expectation is the conventional null value and 2) it has maximum probability of producing the observed data. The rule is employed to develop a complete set of tests for nominal data, and both a one-sample and a two-sample test for difference of means. Numerical illustrations are included for most of the tests developed.

TABLE OF CONTENTS

		Page
1.	Characterization of Hypothesis Testing as Bayesian Inference	4
2.	Previous Research on Choice of a Prior	7
3.	Binomial Test	12
4.	Analysis of Contingency Tables	20
5.	One Sample Difference of Means Test	36
6.	Two Sample Difference of Means Test	45
7.	Discussion	52

1. Characterization of Hypothesis Testing as Bayesian Inference

When an experiment is over or a field study is complete, the researcher must sit himself down and attempt to make some sense out of the mass of data he has collected. Usually he employs the data to judge the credibility of relations among constructs. More often he sifts through the data in the hope of finding some relations. But in either case his inference task boils down to finding an appropriate sampling model and then to computing the probabilities of the sample outcomes predicted by his hypotheses.

Up to this point, the steps in the analysis are pretty much the same both for those who look at the data in "either-or" terms and for those who feel they can distinguish varying shades of gray. Most psychologists are in the "either-or" camp -- either they can reject the null hypothesis (support the research hypothesis) or they can't. When a statistically significant relationship is found, the researcher joyfully calls a halt to celebrate his good fortune. Very seldom does he try to estimate the strength of the relation even though some measures do exist.

The intent of this paper is to propose a very conservative way for assessing the strength of a relation. A natural way to approach this problem is to translate the research hypothesis into a statement about some output parameter. Thus, the stronger the hypothesized relation, the larger the output parameter. Since the parameter is a random variable, it must be described in probabilistic terms, i.e., in terms of a mean, a variance, and the probability that the true value of the parameter falls in a given range.

Let's illustrate these notions with a couple of examples. Suppose we're studying the relationship between union activity and agreement with the union's political line (c.f. Wilensky, 1956). We dichotomize each variable and then cast the data into the following contingency table,

	High Activity	Low Activity	
Agreement with political line	n_{11} $\bar{n}_{11}$	n_{12} $\bar{n}_{12}$	$n_1 = n_{11} + n_{12}$
Disagreement with political line	n_{21} $\bar{n}_{21}$	n_{22} $\bar{n}_{22}$	$n_2 = n_{21} + n_{22}$
			$n = n_1 + n_2$

where n_{ij} and $\bar{n}_{ij}$ are the actual and expected frequencies, respectively. Instead of asking, "Is there a significant relationship between union activity and agreement with political line," we prefer to pose questions like, "What is the probability that,

$$0.2 < \left(\frac{n_{11}}{n_{11}+n_{12}} - \frac{n_{21}}{n_{21}+n_{22}}\right) < 0.4$$

i.e., that the difference in proportions between highly active agreers and highly active disagreers lies between 0.2 and 0.4?"

Or suppose we're studying the relative effectiveness of two job training programs, on-the-job training versus vocational training in a special school. We select two matched samples, assign individuals randomly to the two programs, and observe the difference in annual income between pair members one year after training begins. As advisors to policy makers

we're very interested both in the average value of the income difference and in the probability that the difference is at least $1000.

The law of conditional probability provides a natural framework for answering these kinds of questions. Denote the output parameter by y and the sample data by the vector $\underline{x} = (x_1, x_2, \ldots . x_n)$. Then the inference problem can be transformed into computing the probabilistic description of y, given $\underline{x}$. Typically we describe a random variable in terms of the probability that it lies in a small interval, say $[y_0, y_0 + dy_0]$. Thus,

$$\Pr\{y_0 \le y \le y_0 + dy_0 \mid \underline{x}\} = \frac{\Pr\{\underline{x} \mid y_0 \le y \le y_0 + dy_0\} \Pr\{y_0 \le y \le y_0 + dy_0\}}{\Pr\{\underline{x}\}} \qquad 1]$$

If we define the probability density functions

$$f_y(y_0)\, dy_0 \equiv \Pr\{y_0 \le y \le y_0 + dy_0\}$$

$$f_{y|x}(y_0|x) \equiv \Pr\{y_0 \le y \le y_0 + dy_0 \mid x\},$$

equation 1] may be rewritten as the continuous form of Bayes Law:

$$f_{y|\underline{x}}(y_0|\underline{x}) = \frac{\Pr\{\underline{x}|y_0\} f_y(y_0)}{\int \Pr\{\underline{x}|y_0\} f_y(y_0)\, dy_0} \qquad 2]$$

The first factor in the numerator, $\Pr\{\underline{x} \mid y_0\}$ is merely the sampling model, which gives the probability of observing the data $\underline{x}$ if the true value of the output parameter is y_0 . The second factor represents our prior feelings about y, that is, what we knew about y before the data were collected. The denominator is the usual normalizing factor, $\Pr\{\underline{x}\}$, expressed as a sum over the set of mutually exclusive and collectively exhaustive joint events: the data are $\underline{x}$ and $y_0 \le y \le y_0 + dy_0$.

The mathematical form of the sampling model is objectively determined by the way the data are categorized and by the relationships among the variables in the research hypothesis. However, there is presently no general agreement on a procedure for selecting a prior distribution. We must devise such a procedure if we are to fulfill the promise made above -- develop a conservative way to estimate the strength of a relationship. Before proposing our own solution to this problem, let's pause briefly to review previous suggestions for choice of a prior.

2. Previous Research on Choice of a Prior

Raiffa and Schlaifer (1961, Ch. 3) note that the choice of a prior is simple enough when a body of prior data exist. One merely selects a functional form rich enough to accommodate a wide range of posterior density functions (e.g., skewed, flat, multimodal, etc.), and then chooses prior parameters that best fit the data.

When the researcher lacks either data or firm opinions, choice of a prior is difficult to make and more difficult to defend. For example,

in estimating the parameter in a Bernoulli process, one can characterize the process by $\pi \equiv ln\left(\frac{p}{1-p}\right)$ just as well as by p itself. Clearly a uniform prior on p does not imply a uniform prior on π.

Raiffa and Schlaifer (1961, p. 66) conclude that

> The notion that a broad distribution is an expression of vague opinions is simply untenable . . . Notice, however, that although we cannot distinguish meaningfully between "vague" and "definite" prior distributions, we can distinguish meaningfully between prior distributions which can be substantially modified by a small number of sample observations from a particular data generating process and those which cannot. (sic)

By sensitivity analysis, one can define a set of priors which have this "open-minded" property.

Edwards, Lindman and Savage (1963) outline a procedure called stable estimation, for which the assumption of a uniform prior permits a very good approximation to the actual posterior distribution. Consider an arbitrary interval $[y_1, y_2]$ in the parameter being estimated. Stable estimation can be used when

1) $[y_1, y_2]$ is small and highly favored by the data;
2) the actual prior density changes little in $[y_1, y_2]$; and
3) nowhere outside $[y_1, y_2]$ is the actual prior extraordinarily large compared with its value inside the interval.

If these conditions hold, the posterior is merely the normalized, conditional sampling distribution, whose propoerties can be calculated easily.

Although Edwards et al. briefly discuss prior-posterior analysis and estimation, their treatment of hypothesis testing follows the classical

decision theoretic notion of using data D to choose between a null and an alternative nypothesis, denoted by H_0 and H_1, respectively. Noting that

$$\Pr\{H_0 | D\} = \Pr\{D | H_0\} \Pr\{H_0\} / \Pr\{D\}$$

$$\Pr\{H_1 | D\} = \Pr\{D | H_1\} \Pr\{H_1\} / \Pr\{D\}$$

one may state the test in the form of an odds equation:

$$\Omega(H_0, H_1 | D) = L(H_0, H_1; D)\, \Omega(H_0, H_1) \qquad 3]$$

where

$$\Omega(H_0, H_1) \equiv \Pr\{H_0\} / \Pr\{H_1\}$$

$$L(H_0, H_1; D) \equiv \Pr\{D | H_0\} / \Pr\{D | H_1\}$$

In general H_0 and H_1 can be diffuse. Let λ be the parameter of interest and $f_0(\lambda)$, $f_1(\lambda)$ be the probability density functions corresponding to H_0 and H_1, respectively. Then

$$L(H_0, H_1; D) = \frac{\int \Pr\{D | \lambda\} f_0(\lambda)\, d\lambda}{\int \Pr\{D | \lambda\} f_1(\lambda)\, d\lambda} \qquad 3a.$$

Thus, the likelihood ratio, $L(H_0, H_1; D)$ is merely the ratio of the unconditional sampling distributions corresponding to H_0 and H_1. In many cases it is possible to find bounds for the likelihood function such that any value within the bounded interval completely overwhelms

reasonable choices for the prior odds. Edwards et al. devote the bulk of their analysis to simplifying 3a] in order to generate these bounds. They propose no objective way to determine either the density functions f_0, f_1 or the prior odds, $\Omega(H_0, H_1)$.

One of the few published attempts to apply this theory is Ditz's (1968) analysis of two experiments on the perception of rotary motion. We shall compare his analysis with our own in a later section.

A very different approach to the choice of a prior involves the concepts of entropy and testable information. Suppose one is estimating a parameter y. Testable information about y constrains the form of the prior. Jaynes (1968, p. 229) gives some examples:

1) $y < 6$;

2) the mean value of $\tanh^{-1}(1-y^2)$ in previous measurements was 1.37 ;

3) there is at least a 90 percent probability that $y > 10$.

A meaningful prior distribution for y is the one which maximizes entropy while satisfying the testable constraints. Good (1965, p. 73) reports that

> for a two-dimensional population contingency table with assigned marginal totals, the principle [of maximum entropy] leads to the null hypothesis of no association. A similar conclusion applies to an m-dimensional table. . . In general we could describe the effect of the principle . . . by saying that, in some sense it pulls out the hypothesis in which the amount of independence is as large as possible.

Jaynes uses the principle to generate several discrete probability distributions but notes that the continuous case is troublesome because of the difficulty in defining an invariant measure of the entropy. He obtains continuous priors by <u>ad hoc</u> invariance arguments, which are based

upon the equivalence of probabilities seen by different observers. Since continuous priors are the only ones of interest to us, we cannot use the principle of maximum entropy until the continuous case is solved by a well defined procedure.

A third way of defining a prior relies on extracting information from the sampling distribution. In an analysis of a normal process with an unknown mean and a normal prior, Clutton-Brock (1965) chooses the prior parameter in order to maximize the likelihood, i.e., the unconditional sampling distribution. Here, at last, is something on which to build.

Our procedure for determining a prior borrows ideas from both Jaynes and Clutton-Brock. In a word, we suggest that a very conservative prior can be obtained by 1) constraining the prior mean to lie at the conventional null point, and then 2) selecting prior parameters which maximize the likelihood.

Such a prior has two interesting features. First, because we view hypothesis testing as parameter estimation, the argument of the prior is a continuous random variable. Thus, we have discarded the "either-or ," point null in favord of a "gray" diffuse null, whose expectation is the conventional null value of the parameter. Second, by choosing the prior parameters to max the constrained likelihood, we effectively define a null-centered prior which has maximum probability of generating the observed sample data. For a given functional form, the resulting prior stacks the cards against the research hypothesis as much as possible. Consequently, estimation based on this prior amounts to an *a fortiori* "test" of the experimental effect.

In the next four sections we apply these notions to binomial sampling, analysis of contingency tables, normal sampling with unknown mean and

variance, and sampling from two different normal populations. For analytical convenience, conjugate priors are used throughout; however, the method can be used with any distribution which has a finite mean. Numerical comparisons with max likelihood and conventional tests are included in the development which follows.

3. Binomial Test

In the course of binomial sampling from a population with unknown Bernoulli parameter p, we observe r successes in n trials, that is, r cases fall in one nominal category and n-r cases, in the other. Moreover, let us suppose that the conventaional null value for p is p_0 (0.5 is a very common value for p_0).

The conditional sampling distribution is the familiar binomial expression:

$$\begin{aligned} P_r\{\text{Data} \mid p\} &= P_r\{r \mid p, n\} \\ &= \binom{n}{r} p^r (1-p)^{n-r} \end{aligned} \qquad 4]$$

The beta function is chosen as the prior on p both because a beta has a flexible form and because it combines easily with the conditional sampling distribution. A function with the second property is called a conjugate prior. Denote the prior on p by the symbol $f'(p)$ and restrict the prior to be unimodal. Then

$$\begin{aligned} f'(p) &= f_\beta(p \mid r', n') \\ &= C\, p^{r'} (1-p)^{n'-r'-1} \; ; \quad n' \geq 2 \, , \; n' > r' \end{aligned} \qquad 5]$$

where

$$\mathbb{C}^{-1} = B(r', n'-r')$$

$$B(a,b) \equiv \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \qquad 6]$$

and r',n' are the prior parameters, which must be determined.

The unconditional probability of obtaining the data is merely

$$P_r\{Data\} \equiv \mathbb{P}(r,n\,|\,r';n') = \int_0^1 P_r\{Data\,|\,p\}\,f'(p)\,dp$$

$$= \frac{\Gamma(n')}{\Gamma(r')\Gamma(n'-r')}\;\frac{\Gamma(r+r')\Gamma(n+n'-r-r')}{\Gamma(n+n')}\;\frac{\Gamma(n+1)}{\Gamma(r+1)\Gamma(n-r+1)} \qquad 7]$$

which is a beta-binomial distribution on r. Since the prior mean must fall at the conventional null point, we require that

$$r'/n' = p_0 \qquad 8]$$

and thereby reduce 7] to an equation in one unknown:

$$\mathbb{P}(n'\,|\,r,n,p_0) = \frac{\Gamma(n')}{\Gamma(p_0 n')\Gamma[(1-p_0)n']}\;\frac{\Gamma(p_0 n'+r)\Gamma[(1-p_0)n'+n-r]}{\Gamma(n+n')}\;\frac{\Gamma(n+1)}{\Gamma(r+1)\Gamma(n-r+1)} \qquad 9]$$

In order to select that prior which has maximum probability of producing the data, we

$$\max_{p_0 n' > 1} \mathbb{P}(n'\,|\,r,n,p_0) \qquad 10]$$

But $\mathcal{P} \equiv \ln \mathbb{P}$ will have its maximum at the same n' as $\mathbb{P}$.

Thus 11] may be rewritten as

$$\max_{P_o n' > 1} \mathcal{P}(n' \mid r, n, P_o) \qquad 11]$$

When $\hat{n}'$, the maximizing value of n', is an interior point, its value may be obtained by setting the derivative of $\mathcal{P}$ equal to zero. Thus, $\hat{n}'$ must satisfy the equation

$$0 = \{\psi(\hat{n}') - \psi(\hat{n}'+n)\} - P_o\{\psi(\hat{n}'P_o) - \psi(\hat{n}'P_o+r)\}$$

$$- (1-P_o)\{\psi[(1-P_o)\hat{n}'] - \psi[(1-P_o)\hat{n}'+n-r]\} , \qquad 12]$$

subject to the constraint

$$\hat{r}' = P_o\hat{n}' \geq 1$$

and where $$\psi(x) \equiv \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)} \qquad 13a,b]$$

Equation 12] must be solved numericallly because both the gamma function and its derivative are transcendental functions.

By substituting the series definition for $\psi(x)$ into 12], one can show that there can be only one internal extremum and this extremum must be a maximum. Thus,

$$\hat{n}' = \begin{cases} 2 & \text{if } \frac{d}{dn'}\mathcal{P} \leq 0 \text{ at } n' = 2; \\ \text{Solution to 12] otherwise} & \end{cases} \qquad 14]$$

With the prior parameters now determined, we may solve for $f''(p)$, the posterior on p, and then compute any posterior statistic we choose. According to Bayes Law,

$$f''(p) = \frac{P_r\{\text{Data} \mid p\}\, f'(p)}{\int_0^1 P_r\{\text{Data} \mid p\}\, f'(p)\, dp} = f_\beta(p \mid r'', n'') \qquad 15]$$

where $$r'' \equiv r + r'$$
$$n'' \equiv n + n'$$

This very convenient result flows from the choice of a prior which is conjugate to the conditional sampling distribution.

For comparison with the classical binomial test, we shall compute the posterior tail probability, i.e., the probability that p lies on the improbable side of the null point. This *a fortiori* significance level is

$$\begin{aligned} \alpha_0 \equiv P_r\{p \le p_0\} &= \int_0^{p_0} f''(p)\,dp \\ &= \int_0^{p_0} f_\beta(p \mid r'', n'')\,dp \\ &= I_{p_0}(r'', n''-r'') \end{aligned} \qquad 16]$$

which is a normalized, incomplete beta function.

The posterior mean and variance -- $\bar{p}''$ and $\check{p}''$, respectively-- give a conservative idea of the strength of the effect being studied.

$$\begin{aligned} \bar{p}'' &= r''/n'' \\ \check{p}'' &= \frac{1}{n''+1}\,\bar{p}''(1-\bar{p}'') \end{aligned} \qquad 17a,b]$$

Numerical Examples

To give the reader an idea of the test's conservatism, let's compare its results with some calculations made by Edwards, Lindman, and Savage (1963, p. 225), and with an analysis published by Pitz (1968). Edwards

et al. computed the upper portion of Table 1 in order to demonstrate the large difference between their likelihood ratios and the p-levels obtained from a conventional binomial test. The numerator of each likelihood ratio is based on a point hypothesis located at $p = 0.5$; the denominator of $L(p_0;r,n)$ assumes a uniform hypothesis; and the denominator of L_{min} uses a point hypothesis placed at $p = r/n$, the maximum likelihood estimate of p. The Bayesian p-levels are computed from equations 12], 14], and 16].

TABLE 1

Likelihood Ratios under the Uniform Alternative Prior, Minimum Likelihood Ratios, and Bayesian P-Levels, for Various Values of n and for Values of r Just Significant at the .05 Level

	Experiment Number				
	1	2	3	4	
n	50	100	400	10,000	(very large)
r	32	60	220	5,098	(n + 1.96 n)/2
$L(p_o;r,n)$	.8178	1.092	2.167	11.689	.11689 n
L_{min}	.1372	.1335	.1349	.1465	.1465
prior n	17	33	132	≈ 3500	
prior r	8.5	16.5	66	≈ 1750	
prior $\bar{p}$	.640	.600	.550	.510	
posterior $\bar{p}$	.605	.575	.538	≈ .507	
Bayesian p-level	.042	.041	.041	.042	
Conventional p-level	.050	.050	.050	.050	

Note the close agreement between the Bayesian and the conventional p-levels. Because the experimental effect is very large here, the Bayesian test indicates a stronger result than the conventional binomial test. However, this apparent lack of conservatism disappears if the roles of $r/n \equiv \tilde{p}$ and the null value of p are interchanged in the conventional test. This redefinition of p effectively reduces the standard deviation in the conventional test statistic.

Much more interesting than the p-levels are the values of the posterior mean, $\bar{p}''$. As we would expect, $\bar{p}'' < \bar{p}'$. Note that for values of r/n very near ½, the experimental effect is quite small, according to the conservative Bayesian estimate.

Pitz used Edwards et al.'s likelihood ratio technique to discriminate between two theories about the perception of rotary motion in depth. Day and Power (1965) hypothesized inability to identify the direction of rotation while Hershberger (1967) predicted the opposite and subsequently obtained experimental evidence supporting his theory. Thirty-two of Hershberger's forty-eight subjects reported the direction of rotation correctly, yielding an exact one tail p-level of 0.0154 under the null hypothesis of $p = \frac{1}{2}$ (i.e. the hypothesis that Day and Power's theory is correct).

Pitz computed likelihood ratios for three different priors on p (see Figure 1). According to Pitz (p. 254),

> The assumption of a uniform prior implies that the stimulus cues that form the basis for Hershberger's theory may not be very effective determinants of perception. A strong belief in the effectiveness of these cues might be represented by a probability distribution . . . such as . . . $g_2(p)$ . . . A third distribution, $g_3(p)$, illustrates a still more firm commitment to the idea that p must be large . . . The value of $g_3(2/3)$ was deliberately chosen to be 0.394, since for this value $[L(p_0; r, n)]$ would be 1.0*

* To avoid unnecessary confusion, we have substituted "p" for Pitz's "ϱ" and have used Edwards et al.'s symbol for the stable estimation value

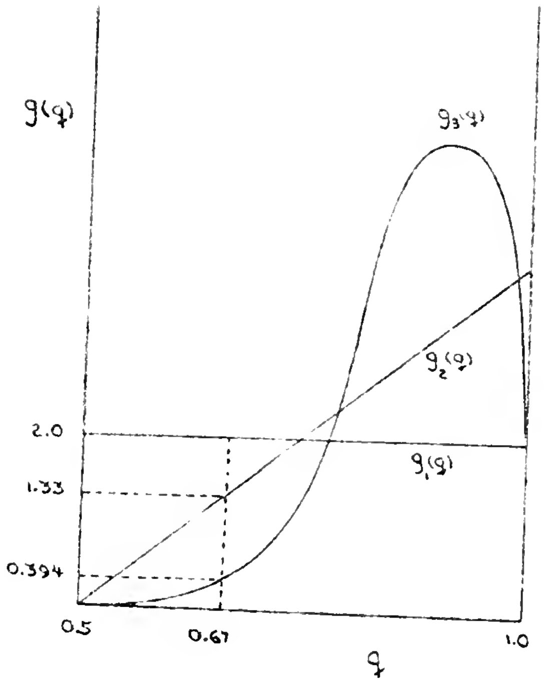


Fig. 1. Three possible prior probability distributions defined for values of Q, given the correctness of Hershberger's theory. (It is assumed that values of Q less than 0.5 are impossible.)

Table 2: Comparison of Pitz's Likelihood Ratio Results with Bayesian and Conventional Binomial Tests

Data: 32 successes in 48 trials

		$L(p_o;r,n)$:		L_{min}	Bayesian p-level	Conventional p-level
prior:	$g_1(p)$	$g_2(p)$	$g_3(p)$			
	0.197	0.296	1.0	0.066	0.0177	0.0154

Table 2 compares Pitz's three values for $L(p_0; r, n)$ with our results for L_{min} and the two binomial tests. The Bayesian p-level is based on a beta prior with parameters n'=11.0, r'=5.5, a mean of 0.5 and a variance of 0.0208. Thus, about 95 % of the prior null hypothesis is contained in the interval [0.36, 0.64]. Either a sharper or a more diffuse null would have lower probability of producing the observed data.

A puzzling feature of both Tables 1 and 2 is the wide discrepancy between the likelihood ratios and the Bayesian p-levels, and the variation among the likelihood ratios themselves. However, one must remember that a Bayesian p-level and a likelihood ratio are two very different beasts: the first <u>is the probability</u> that the parameter of interest lies in a certain portion of the tail of the posterior hypothesis; whereas, the second <u>compares the probabilities</u> that two rival hypotheses produced the data. If the rivals are point hypotheses, e.g., $p = p_0$ and $p = p_1$, and if one hypothesis, say p_0, lies so far out on the tail of the sampling distribution that

$$\int_0^{p_0} f(p)\,dp \doteq f(p_0) ,$$

then the likelihood ratio and Bayesian p-level ought to be at least the same order of magnitude.

But if one hypothesis is diffuse, then the tactic of stable estimation permits one to generate virtually any value of $L(H_1, H_2; D)$ he pleases! As Pitz has demonstrated so well, one merely takes care to sketch an

---

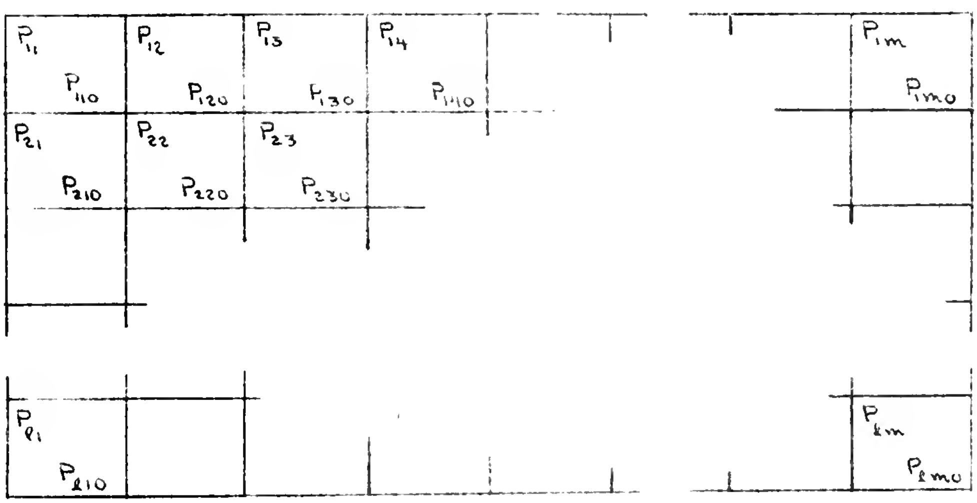
of the likelihood ratio.

hypothesis which has the appropriate value at the mode of the sampling distribution (see Figure 1). Of course, it is not feasible to pick this modal value of the hypothesis before the data are collected. But worse yet, Edwards et al. give us no objective way of picking it even after the data are in hand. The test developed here represents one attempt to put the choice of the prior on an objective basis.

In the next section of the paper we extend the binomial test to the analysis of contingency tables.

4. Analysis of Contingency Tables

Consider an lxm contingency table, with p_{ij} and p_{ijo} being, respectively, the actual and conventional expected values of the population proportion corresponding to the i,j segment:



We feel that the probabilities of interest to the researcher are of the type

$$\Pr\{p_{ij} \ge p_{ij0}\}$$

$$\Pr\{p_{ij} \ge p_{ij0} \text{ and } p_{jl} \ge p_{jl0} \text{ and } p_{jk} \le p_{jk0}\}$$

$$\Pr\{\alpha \le | p_{ij} - p_{ik} | \le \beta\} \qquad 18]$$

etc.

Our goal is to derive expressions 18] for the general m x n table. However, we will work up to this result gradually by first analyzing the 1 x m and 2 x 2 cases.

4.1 The 1 x m Table

Suppose an independent random sample yields r_i events in the i^{th} category, there being n events in all. Define $p_i \equiv r_i/n$ and p_{io} as, respectively, the actual and expected proportion of events of type i. Thus, the contingency table for the data looks like this:

								Totals
Category	1	2	3		i		m	
Actual Frequency	r_1	r_2	r_3		r_i		r_m	n
Actual Proportion	p_1	p_2	p_3		p_i		p_m	1
Expected Proportion	p_{10}	p_{20}	p_{30}		p_{i0}		p_{m0}	1

The conditional sampling distribution is the multinomial:

$$\Pr\{r_1, r_2, r_3, \ldots r_{m-1} \mid n, p_1, p_2, p_3, \ldots p_{m-1}\} = n! \prod_{i=1}^{m} \frac{p_i^{r_i}}{r_i!} \qquad 19]$$

where
$$\sum_{i=1}^{m} p_i = 1$$
$$\sum_{i=1}^{m} r_i = n \qquad 20\ a.b]$$

As for the binomial case, the max likelihood estimate for p_i is merely r_i/n. This is easily shown by substituting 20] into 19] and then differentiating the log of the resulting expression. Using the binomial development as a guide, we select as prior the multivariate beta density with means equal to the conventaionlly defined expected proportions.

$$f'(p_1, p_2, \ldots, p_{m-1} | r_1', r_2', \ldots, r_{m-1}, n) = \prod_{i=1}^{m} \frac{p_i^{r_i'-1}}{\Gamma(r_i')} \qquad 21]$$

where
$$0 \leq p_i \leq 1$$
$$\sum_{i=1}^{m} r_i' = n'$$
$$r_i' = n' p_{io} \geq 1 \qquad 22\ a.b.c]$$

The unconditional sampling distribution, the probability of obtaining the observed, data is obtained as before.

$$Pr\{r_1, r_2, \ldots, r_{m-1} | n\} \equiv \mathbb{P}$$

$$= \int dp_1 \int dp_2 \ldots \int dp_{m-1}\ Pr\{r_1, r_2, \ldots r_{m-1} | n, p_1, p_2, \ldots p_{m-1}\} f'(p_1, p_2, \ldots p_{m-1} | r_1' \ldots r_{m-1}', n')$$

$$= \frac{\Gamma(n+1)\Gamma(n')}{\Gamma(n+n')} \prod_{i=1}^{m} \frac{\Gamma(r_i+r_i')}{\Gamma(r_i')\Gamma(r_i'+1)} \qquad 25]$$

Upon substituting 22c] into 25], taking the natural log, and differentiating with respect to n', we obtain the equation which determines the internal maximizing value of n'.

$$0 = [\psi(n') - \psi(n+n')] - \sum_{i=1}^{m} p_{io}[\psi(p_{io}n') - \psi(p_{io}n'+r_i)] \qquad 26]$$

The global max $\hat{n}'$ is either m or the solution to 26] , whichever yields the larger value of $\mathbb{P}$. As in the 1 x 2 case, 26] is a transcendental equation and must be solved numerically.

Because the multivariate beta prior $f'(p_i|r_i')$ is conjugate to the conditional sampling distribution, the posterior $f''(p_i|r_i',r_i)$ is also multivariate beta. In particular,

$$f''(p_1, p_2, \dots p_{m-1}|r_1, \dots r_m; r_1' \dots r_m') = \Gamma(n'') \prod_{i=1}^{m} \frac{p_i^{r_i''-1}}{\Gamma(r_i'')} \qquad 27]$$

where

$$0 \le p_i \le 1$$
$$r_i'' \equiv r_i + r_i'$$
$$n'' \equiv n + n' \qquad 28\text{ a b,c}]$$

The development of the 1 x m test is completed with a calculation of two event probabilities: 1) $P_r\{p_i \le p_{io}\} \equiv \alpha_{io}$, the probability, say, that the ith proportion actually lies on the indicated side of the null point; and 2) $P_r\{\Delta \le (p_i - p_j)\} \equiv \alpha_{ij}(\Delta)$,

the probability that two proportions differ by at least Δ.

Clearly

$$\alpha_{io} = \int_0^{p_{io}} dp_i \int_0^{1-p_i} dp_k \int_0^{1-p_i-p_k} dp_\ell \cdots \int_0^{A} dp_{m-1} \left[1-\sum_{j=1}^{m-1} p_j\right]^{r_m''} \frac{1}{\Gamma(r_m'')} \prod_{i=1}^{m-1} \left[\frac{p_i^{r_i''-1}}{\Gamma(r_i'')}\right]$$

with $\quad A \equiv 1 - \sum_{j \neq m-1} p_j$ 29]

Although formidable appearing, 29] is easily integrated to yield

$$\alpha_{io} = I_{p_{io}}(r_i'', n'' - r_i'') \qquad 30]$$

where $I_x(a,b)$ is the incomplete beta function:

$$I_x(a,b) \equiv \frac{1}{B(a,b)} \int_0^x u^{a-1}(1-u)^{b-1} du \qquad 31]$$

In order to calculate $\alpha_{ij}(\Delta)$ we begin with the marginal on p_i and p_j. With no loss of generality, we can assign i and j the values 1 and 2. The calculation of α_{io} easily leads us to discover that

$$f(p_1, p_2 | r_1, r_2, n) = C_{12}\, p_1^{r_1-1}\, p_2^{r_2-1}\, (1-p_1-p_2)^{r_3-1}$$

with

$$C_{12}^{-1} \equiv \Gamma(r_1)\Gamma(r_2)\Gamma(r_3)/\Gamma(n)$$

$$r_3 \equiv n - r_1 - r_2$$

32(a,b,c]

Now define

$$\Delta_{21} \equiv P_2 - P_1 \tag{35}$$

and consider the event $\Delta_{21} \leq \Delta$. We must consider two cases: $\Delta \leq 0$ and $\Delta > 0$. The event $\Delta_{21} \leq \Delta$ for a negative Δ is portrayed in Figure 2. Let's analyze this situation first.

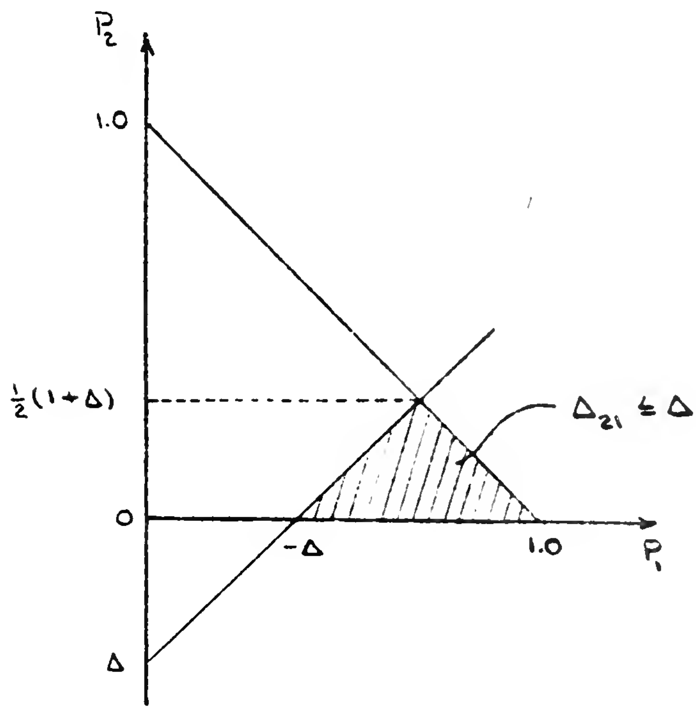


Figure 2: Region of Integration Used in Computing the Cumulative Probability, $P_r\{\Delta_{21} \leq \Delta\}$

Clearly,

$$P_r\{\Delta_{21} \leq \Delta\} = \iint dp_1\, dp_2\, f(p_1, p_2 \mid r_1, r_2, n)$$

$$= \mathbb{C}_{12} \int_0^{\delta_1} dP_2 \int_{P_2-\Delta}^{1-P_2} dP_1\, P_1^{r_1-1} P_2^{r_2-1} (1-P_1-P_2)^{r_3-1}$$

$$= I_{\delta_1}(r_2, n-r_2) - \mathbb{C}_{12} \mathcal{J}_1$$

where

$$\delta_1 \equiv \tfrac{1}{2}(1+\Delta)$$

and

$$\mathcal{J}_1 \equiv \int_0^{\delta_1} dp_2 \int_0^{p_2-\Delta} dp_1\, p_1^{r_1-1} p_2^{r_2-1} (1-p_1-p_2)^{r_3-1} \qquad 34\,a,b,c]$$

After the transformation $x = p_1/(1-p_2)$

the integral $\mathcal{J}_1$ becomes

$$\mathcal{J}_1 = \int_0^{\delta_1} dp_2 \int_0^{\delta_3} dx\, x^{r_1-1}(1-x)^{r_3-1}\, p_2^{r_2-1}(1-p_2)^{r_1+r_3-1}$$

with 35]

$$\delta_3 \equiv (p_2-\Delta)/(1-p_2)$$

If r_1, r_2, r_3 are integers, one may use formula [16.1a)] in Grobner and Hofreiter (1961) to obtain

$$\mathcal{J}_1 = \sum_{\nu=0}^{r_3-1} \left\{ \int_0^{\delta_1} dp_2 (1-p_2)^{\nu}\, p_2^{r_2-1} (p_2-\Delta)^{r_1} (1+\Delta-2p_2)^{r_3-\nu-1} \frac{(r_3-1;-1;\nu)}{(r_1+r_3-1;-1;\nu+1)} \right\}$$

where 36 a,b]

$$(m;d;\nu) \equiv m(m+d)(m+2d)\ldots\ldots[m+(\nu-1)d]$$

Finally, $\mathcal{J}_1$ can be expressed as a sum of hypoergeometric functions, by applying formula [3.2.11] from Gradshtein and Ryzhik (1965, p. 287)

$$\mathcal{J}_1(r_1, r_2, r_3, -\Delta) =$$

$$-2^{r_2} \sum_{\nu=0}^{r_3-1} \left\{ a_\nu (-\Delta)^{r_1} (1+\Delta)^{r_2+r_3-\nu-1} F\left[r_2, -\nu, -r_1, r_2+r_3-\nu; \delta_1, \delta_1/\Delta\right] \right\}$$

37]

where $a_\nu \equiv (r_3 - 1 ; -1 ; \nu) / (r_1 - r_3 - 1 ; -1 ; \nu + 1)$

combining 34 and 37 , we obtain α_{12} for the case of $\Delta \leq 0$.
For the case $\Delta > 0$, one can now show quite easily that

$$\alpha_{12}(\Delta > 0) = 1 - I_{\delta_2}(r_1, n - r_1) + \mathbb{C}_{12}\vartheta_1(r_2, r_1, r_3, \Delta)$$

with
$$\delta_2 \equiv \tfrac{1}{2}(1 - \Delta) \qquad 38a,b]$$

Let us summarize the above calculation of $\alpha_{12}(\Delta)$:

$$\alpha_{12}(\Delta) = \begin{cases} I_{\delta_1}(r_2, n - r_2) - \mathbb{C}_{12}\vartheta_1(r_1, r_2, r_3, -\Delta) & ; \; -1 \leq \Delta \leq 0 \\ 1 - I_{\delta_2}(r_1, n - r_1) - \mathbb{C}_{12}\vartheta_1(r_2, r_1, r_3, \Delta) & ; \; 0 < \Delta \leq 1 \end{cases}$$

39a,b]

where
$$r_3 \equiv n - r_1 - r_2$$

$$\delta_1 \equiv (1 + \Delta)/2$$

$$\delta_2 \equiv (1 - \Delta)/2$$

$$\mathbb{C}_{12} \equiv \Gamma(r_1)\Gamma(r_2)\Gamma(r_3)/\Gamma(n)$$

and ϑ_1 , is defined by equations 37] and 36b] .

<u>Example</u>

A trichotomy with 15 observations provides a simple numerical application fo the 1 x m test:

	1	2	3	Totals
Category	1	2	3	
Actual Frequency	1	4	10	15
Expected Frequency	5	5	5	15

Because the formulae for $\alpha_{ij}(\Delta)$ are rather complicated, we shall limit the analysis to calculation of the α_{i0}.

For these data, the global max of the unconditional sampling distribution occurs at the boundary points n' = 3. The posterior parameters and individual "tail probabilities" can be immediately computed from 22] and 30]. Table 3 compares these Bayesian results with a conventional χ^2 analysis.

Bayesian Analysis				Conventional Analysis
Category	1	2	3	
Prior Parameter	1	1	1	$\chi^2(2df) = 8.4$
Posterior Parameter	2	5	11	$P_{\text{one tail}} = 0.0075$
$Pr\{p_i \leq 1/3\}$	0.990	0.719	0.0074	

Table 3: Comparison of Bayesian and Conventional Analysis of a 1 x 3 Contingency Table

Note that the conventional test tells one only that the data, taken as a whole, constitute a compound rare event. In contrast, the Bayesian analogue looks at each cell and provides 1) a tail probability which is maximally prejudiced against the data, and 2) a conservative mean value for each proportion.

Of course, we could compute a global statistic comparable to the conventional χ^2. It would be merely the probability of the compound event,

$$\{ P_1 \leq P_{10} , P_2 \geq P_{20} , \ldots\ldots\ldots P_m \leq P_{m0} \} ,$$

where the inequalities are chosen to represent the case exemplified by the observed data. But why do it? We are really interested in knowing how much confidence to place in the <u>detailed structure of the data.</u> The Bayesian test gives conservative answers to queries about fine structure; whereas, the conventional approach affords such answers (although rarely required) by the clumsy technique of binomial tests.

4.2 The 2 x 2 Table

The analysis of a 2 x 2 contingency table is the next step toward the treatment of the ℓ x m case. Suppose we have drawn independent random samples from populations I and II, and then categorized each observation as "A" or "not-A". Such data are usually presented in a 2 x 2 table:

		Attribute A	Attribute not-A	
Population	I	r_1	$n_1 - r_1$	n_1
Population	II	r_2	$n_2 - r_2$	n_2
		$r_1 + r_2$	$n - r_1 - r_2$	n

where r_1, r_2 ~ observed frequencies of type A in populations I and II, respectively;

n_1, n_2 ~ size of sample from populations I and II, respectively.

Suppose p_1 and p_2 are the true proportion of type A's in the two populations. Then the conditional sampling distribution is the product of two binomials.

$$P_r\{r_1, r_2 \mid n_1, n_2; p_1, p_2\} = \binom{n_1}{r_1} p_1^{r_1} (1-p_1)^{n_1-r_1} \binom{n_2}{r_2} p_2^{r_2} (1-p_2)^{n_2-r_2} \qquad 40]$$

A convenient conjugate prior is the matrix beta :

$$f'(p_1, p_2) = f_{M-B}(p_1, p_2 \mid r_1', n_1', r_2', n_2') = \mathbb{C}\, p_1^{r_1'-1} (1-p_1)^{n_1'-r_1'-1}\, p_2^{r_2'} (1-p_2)^{n_2'-r_2'-1}$$

where $\qquad 41]$

$$\mathbb{C}^{-1} = B(r_1', n_1'-r_1')\, B(r_2', n_2'-r_2')$$

$$0 \le p_i \le 1$$

$$1 \le r_i' \le n_i' \quad , \quad (i = 1, 2) \qquad 42a,b,c]$$

And the unconditional sampling distribution $\mathbb{P}$ is computed in the usual way.

$$\mathbb{P} = \mathbb{C} \int_0^1 dp_1\, p_1^{r_1'-1} (1-p_1)^{n_1'-r_1'-1} \binom{n_1}{r_1} p_1^{r_1} (1-p_1)^{n_1-r_1} \int_0^1 dp_2\, p_2^{r_2'-1} (1-p_2)^{n_2'-r_2'-1} \binom{n_2}{r_2} p_2^{r_2} (1-p_2)^{r_2-r_2}$$

$$= \binom{n_1}{r_1} \frac{B(r_1'', n_1''-r_1'')}{B(r_1', n_1'-r_1')} \cdot \binom{n_2}{r_2} \frac{B(r_2'', n_2''-r_2'')}{B(r_2', n_2'-r_2')} \qquad 43]$$

where

$$r_i'' = r_i' + r_i$$

$$n_i'' = n_i' + n_i \qquad 44]$$

In order to determine internal extrema of 43] , we set $d(\ln \mathbb{P}) = 0$ and so obtain the two independent conditions,

$$0 = [\psi(n'_i) - \psi(n'_i + n_i)] - p_{io}[\psi(p_{io} n'_i) - \psi(p_{io} n'_i + r_i)]$$

$$- (1-p_{io})\{\psi[(1-p_{io}) n'_i] - \psi[(1-p_{io}) n'_i + n_i - r_i]\}, \quad (i = 1, 2) \qquad 45]$$

where $p_{io} = r'_i / n'_i = \frac{n_i}{n}(r_1 + r_2)$, $\qquad 46]$

in accord with the rule of equating prior means with conventional null values.

Equations 45] have precisely the same form as 12] , the condition determining the internal max in the binomial test. Thus, we immediately conclude that the global max for $\mathbb{P}$ occurs at the point $(\hat{n}'_1, \hat{n}'_2)$, where

$$\hat{n}'_i = \begin{cases} 2 & \text{if } \frac{d}{dn'_i} \ln \mathbb{P} \leq 0 \quad \text{at } n'_i = 2 \\ \text{solution to 45] otherwise} \end{cases}$$

Because the matrix beta prior is conjugate to the conditional sampling distribution, the posterior is also matrix beta, that is,

$$f''(p_1, p_2 \mid \text{Data}) = f_{M\text{-}B}(p_1, p_2 \mid r''_1, n''_1, r''_2, n''_2) \qquad 47]$$

with posterior parameters defined by equations 44] above.

The statistic of interest to the researcher is

$$A_{12} = p_1 - p_2 ,$$

the difference in the proportions of A's in the two populations. It is a straightforward matter to compute $f_{\Delta_{12}}(\Delta)$, the probability density, and $P_{\Delta_{12} \leq}(\Delta)$, the cumulative distribution, for this

difference if r_i', n_i' are restricted to integral values:

$$f_{\Delta_{12}}(\Delta) = \begin{cases} \mathbb{C}\, B(r_2, n_1-r_1)(-\Delta)^{r_1-1}(1+\Delta)^{n_1-r_1+r_2-1} F_1\left[r_2, r_2-n_2+1, 1-r_1, n_1-r_1+r_2 : \frac{1+\Delta}{\Delta} ; 1+\Delta\right] \\ \qquad\qquad (-1 \le \Delta \le 0) \\ -\mathbb{C}\, B(r_1, n_2-r_2)(\Delta)^{r_2-1}(1-\Delta)^{n_2-r_2+r_1-1} F_1\left[r_1, r_1-n_1+1, 1-r_2, n_2-r_2+r_1 : \frac{1-\Delta}{\Delta} ; 1-\Delta\right] \\ \qquad\qquad (0 < \Delta \le 1) \end{cases}$$

$$P_{\Delta_{12}\le}(\Delta) = \tag{48}$$

$$\begin{cases} -\mathbb{C}(1+\Delta)^{r_2}(-\Delta)^{r_1} \sum_{\nu=0}^{n_1-r_1-1} \left\{ a_\nu (1+\Delta)^{n_1-r_1-\nu-1} B(n_1-r_1-\nu, r_2) F_1\left[r_2, -r_1, r_2-n_2+1, n_1-r_1+r_2-\nu : \frac{1+\Delta}{\Delta} : 1+\Delta\right]\right\} \\ + I_{1+\Delta}(r_2, n_2-r_2), \qquad (-1 \le \Delta \le 0) \\ \mathbb{C}(1-\Delta)^{r_1}(\Delta)^{r_2} \sum_{\nu=0}^{n_2-r_2-1} \left\{ b_\nu (1-\Delta)^{n_2-r_2-\nu-1} B(n_2-r_2-\nu, r_1) F_1\left[r_1, -r_2, r_1-n_1+1, n_2-r_2+r_1-\nu : \frac{1-\Delta}{\Delta}, 1-\Delta\right]\right\} \\ + 1 - I_{1-\Delta}(r_1, n_1-r_1), \qquad (0 < \Delta \le 1) \end{cases} \tag{49}$$

where

$$a_\nu \equiv (n_1-r_1-1; -1; \nu) / (n_1-1; -1; \nu+1)$$

$$b_\nu \equiv (n_2-r_2-1; -1; \nu) / (n_2-1; -1; \nu+1)$$

$$\mathbb{C}^{-1} \equiv B(r_1, n_1-r_1)\, B(r_2, n_2-r_2) \tag{50 a, b, c}$$

To avoid needless clutter, primes were omitted in 48] - 50] .

The expression for the cumulative is particularly simply when $\Delta = 0$. Moreover, for this special case, it is possible to get a closed form expression for non-integral values of r_i and n_i (we are still omitting all primes):

$$P_{\Delta_{12}\leq}(\Delta) = \frac{1}{r_2}\, C\, B(r_1+r_2, n_1-r_1)\; {}_3F_2(r_2, 1+r_2-n_2, r_1+r_2; r_2+1, n_1+r_2; 1)$$

$$= \frac{1}{r_2} \frac{B(r_1+r_2, n_1-r_1)}{B(r_1, n_1-r_1)\, B(r_2, n_2-r_2)} \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\Gamma(r_2+k)}{\Gamma(r_2)} \frac{\Gamma(1+r_2-n_2+k)}{\Gamma(1+r_2-n_2)} \frac{\Gamma(r_1+r_2+k)}{\Gamma(r_1+r_2)} \frac{\Gamma(r_2+1)}{\Gamma(r_2+1+k)} \frac{\Gamma(n_1+r_1)}{\Gamma(n_1+r_1+k)} \qquad 51]$$

This looks like a mess, but it really isn't so bad because it will always be possible to choose n_2 and r_2 such that $1+r_2-n_2 < 0$, thereby causing the series to terminate after n_2-r_2-1 terms. We attempted to find a simpler expression for ${}_3F_2$ but had no success.

If n_i, r_i are integral, 51] can be simplified somewhat:

$$P_{\Delta_{12}\leq}(0) = 1 - \left\{ \sum_{\nu=0}^{n_1-r_1-1} a_\nu\, B(r_1+r_2, n-r_1-r_2-\nu-1) \right\} \Big/ \left\{ B(r_1, n_1-r_1)\, B(r_2, n_2-r_2) \right\}$$

where

$$a_\nu \equiv (n_1-r_1-1; -1; \nu) / (n_1-1; -1; \nu+1) \qquad 52]$$

and r_i, n_i are integers.

This completes the development of the 2 x 2 test. Before extending the analysis to an m x n table, let's pause to apply the 2 x 2 results to a numerical example.

Example

Suppose a study of attribute A in two populations yields the following data:

Population	Attribute: A	Attribute: not-A	
I	15 (10)	5 (10)	20
II	5 (10)	15 (10)	20
	20	20	40

Table 4 compares a conventional analysis with a Bayesian analysis in which prior parameters are restricted to integral values. Calculation of other posterior statistics, for example,

Bayesian Analysis	Conventional Analysis
Prior parameters: $r_i' = 2$, $n_i' = 4$	$\chi^2(1\ df) = 10$
Posterior parameters: $r_i'' = 17$, $n_i'' = 24$	
$\Pr\{p_1 - p_2 < 0\} = 0.00138$	$P_{one\text{-}tail} = 0.00078$

Table 4: Comparison of Bayesian and Conventional Analysis of a 2 x 2 Contingency Table

the mean difference in proportions, is omitted; however, the requisite expressions may be found in 48] - 50] above.

4.3 The ℓ x m Table

The results of sections 4.1 and 4.2 are easily extended to an ℓ x m test. The conditional sampling distribution is a product of 1 multinomials, each having m-1 independent proportions. But more important, the prior-posterior family is an ℓ - (m-1) matrix beta of the form

$$f_{M-B}\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1,m-1} \\ \vdots & \vdots & & \vdots \\ P_{\ell 1} & P_{\ell 2} & \cdots & P_{\ell,m} \end{bmatrix} = \prod_{k=1}^{\ell}\left\{ \Gamma(n_k) \prod_{j=1}^{m} \frac{P_{kj}^{r_{kj}-1}}{\Gamma(r_{kj})} \right\} \qquad 53]$$

where

$$n_k \equiv \sum_{j=1}^{m} r_{kj}$$

$$1 = \sum_{j=1}^{m} P_{kj}$$

As in the 1 x m case, we are interested in the probabilities

$$P_r\{ P_{kj} \leq P_{kjo} \}, \qquad 54]$$

P_{kjo} being the null or conventional expected value. But it is easy to see that the expression for 54] is the same as the corresponding quantity in the 1 x m analysis. We need merely insert a subscript in the 1 x m result. Moreover, the probabilities

$$P_r\{ P_{kj} - P_{ki} \leq \Delta \} \qquad 55a]$$

have the same form as the 1 x m results, and the probabilities

$$P_r\{ P_{ki} - P_{ji} \leq \Delta \} \qquad 55b]$$

have the same form as the expression just derived for the 2 x 2 case. Thus, we now have in hand a complete Bayesian counterpart of the classical analysis of nominal data. In the two sections which follow, we use these same notions of conservative estimation to develop a one sample and a two sample test for normal populations.

5. <u>One Sample, Difference of Means Test</u>

Suppose one has obtained n samples $(x_1, x_2, \ldots x_n)$ from a normal population with unknown mean and variance, μ and σ^2, respectively. The conditional sampling distribution is,

$$\Pr\{\underline{x} \mid \mu, \sigma^2\} = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{h}{2} \sum_i (x_i - \mu)^2\right\} \qquad 56]$$

where we have used the precision $h \equiv 1/\sigma^2$ instead of the variance. By differentiating the logarithm of 56] , it is easy to show that the maximum likelihood estimators of μ and h, are merely the sample mean and sample precision, that is,

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \equiv m$$

$$\hat{h} = n \Big/ \sum_{i=1}^{n} (x_i - m)^2 = \frac{n}{n-1} \frac{1}{V} \qquad 57\,a,b]$$

Since $\hat{\mu}$ and $\hat{h}$ involve all of the information necessary to estimate the true mean and precision, we choose the sample mean and precision as the sufficient statistics which must appear in the prior on μ and h.

Raiffa and Schlaiffer (1961, pp. 298-303) note that a normal-gamma prior is conjugate to 56] when both the mean and variance are unknown. Thus, if we choose a prior of the form,

$$f'(\mu, h \mid m', n', \nu', V') = f_{N1}(\mu \mid m', n'h)\, f_{\gamma 2}(h \mid V', \nu'), \qquad 58]$$

the posterior is also normal-gamma, with parameters m'', n'', ν'', V'' which are simple functions of the prior parameters and the data:

$$n'' \equiv n + n'$$

$$\nu'' \equiv \nu + \nu' + 1$$

$$\nu \equiv n - 1; \qquad \nu' \equiv n' - 1$$

$$m'' \equiv (mn + n'm')/n''$$

$$V'' \equiv \left[\nu V + \nu' V' + \frac{nn'}{n''}(m - m')^2\right] / \nu'' \qquad 59]$$

To apply our method, we begin with the unconditional sampling distribution:

$$\mathbb{P} \equiv P_r\{m, V \mid n\}$$

$$= \int_{-\infty}^{\infty} d\mu \int_{0}^{\infty} dh\, f_{N1}(\mu \mid m', n'h)\, f_{\gamma 2}(h \mid V', \nu') \left(\frac{h}{2\pi}\right)^{n/2} \exp\left\{-\frac{h}{2}\sum_i (x_i - \mu)^2\right\}$$

$$= (2\pi)^{-n/2} \left(\frac{n'}{n''}\right)^{1/2} \frac{\Gamma\left(\frac{\nu''}{2}\right)}{\Gamma\left(\frac{\nu'}{2}\right)} \frac{\left(\frac{\nu' V'}{2}\right)^{\nu'/2}}{\left(\frac{\nu'' V''}{2}\right)^{\nu''/2}} \qquad 60]$$

Now we choose that member of the normal-gamma family which maximizes 60 subject to the constraints

$$n' > 1$$

$$E(\mu) = m' = 0 \qquad 61a,b]$$

The first constraint is imposed by the choice of prior, whereas, the second corresponds to the conventional null hypothesis of "no mean difference." In view of 61b] , the expressions for the posterior mean and variance become

$$m'' = \frac{mn}{n+n'}$$

$$V'' = \left[\nu V + \nu' V' + \frac{nn'}{n''} \right] / \nu'' \qquad 62a,b]$$

Upon equating to zero the two partial derivatives of $\mathbb{P}$, we obtain the conditions for an unconstrained internal extremum:

$$\frac{\partial}{\partial V'} \ln \mathbb{P} = 0 = \frac{\nu'}{V'} - \frac{\nu''}{V''} \frac{\partial V''}{\partial V'} \qquad 63a]$$

$$\frac{\partial}{\partial n'} \ln \mathbb{P} = 0 = \frac{1}{n'} - \frac{1}{n''} + \left[\psi\left(\frac{n+n'-1}{2}\right) - \psi\left(\frac{n'-1}{2}\right) \right] + \ln\left(\frac{n'-1}{n+n'-1}\right)$$

$$+ \ln \frac{V'}{V''} - \frac{n+n'-1}{V''} \frac{\partial V''}{\partial n'} \qquad 63b]$$

Equation 63a] yields the very interesting result

$$V' = V'' \qquad 64]$$

In other words, the prior parameters are chosen in such as way that the

posterior variance does not decrease from its prior value. Moreover, when 64] is combined with 62b] to yield,

$$V' = \frac{n-1}{n} V - \frac{n'}{n'} m^2 , \qquad 65]$$

we note that the prior variance generally exceeds the sample variance. Thus, this particular choice of prior parameters will usually not add to the precision of our estimates. Apparently, this result depends upon the functional form of the prior; a gamma-1 prior yields a much more complex relation between V' and V".

After using 62b] and 64] to compute $\frac{\partial V''}{\partial n'}$, we may combine 63b] and 65] to obtain a single condition for constrained internal extrema:

$$0 = \frac{1}{n'} \frac{n}{n+n'} + \left[\psi\left(\frac{n+n'-1}{2}\right) - \psi\left(\frac{n'-1}{2}\right) \right] + \ln\left(\frac{n'-1}{n+n'-1}\right) - \left(\frac{n}{n+n'}\right)^2 \frac{m^2}{V'} \qquad 66]$$

where V' is defined by 65] above. Thus, the maximizing value of n' is defined by the expression,

$$\hat{n}' = \begin{cases} 1 & \text{when } \frac{\partial \mathbb{P}}{\partial n'} \leq 0 \text{ at } n' = 1 \\ \text{the solution to 66] otherwise} & \end{cases} \qquad 67]$$

The one sample test is completed by calculation of certain posterior statistics of interest. An examination of 58] reveals that the marginal distribution on h is $f_{\gamma 2}(h|V'', \nu'')$. Whence it is easy to calculate the posterior mean and variance of h:

$$\bar{h}'' = 1/V''$$

$$\check{h}'' = 2/\nu'' V''^2 \qquad 68a,b]$$

After a little manipulation, one can show that the marginal on μ is a general Student density function, i.e.,

$$f(\mu \mid m'', n'', \nu'', V'') = f_S(\mu \mid m'', \tfrac{n''}{V''}, \nu'')$$

$$= \left(\frac{n''}{V''}\right)^{1/2} \frac{\nu''^{\nu''/2}}{B(\frac{\nu''}{2}, \frac{1}{2})} \left[\frac{n''}{V''}(\mu - m'')^2 + \nu'' \right]^{-\frac{\nu''+1}{2}} \qquad 69]$$

whose mean and variance are, respectively,

$$\bar{\mu}'' = m'' , \qquad \nu'' > 1$$

$$\check{\mu}'' = \frac{V''}{n''} \frac{\nu''}{\nu'' - 2} , \qquad \nu'' > 2 \qquad 70a,b]$$

Finally, the Bayesian tail probability for μ is merely

$$\alpha \equiv P_r \{ \mu \le 0 \}$$

$$= \int_{-\infty}^{0} f_S(\mu \mid m'', \tfrac{n''}{V''}, \nu'') \, d\mu$$

$$= \frac{1}{2} \left\{ 1 - A\left[m'' \left(\frac{n''}{V''}\right)^{1/2} \Big| \, \nu'' \right] \right\}$$

where $A(t|\nu)$ is the tabulated student integral for degrees of freedom:

$$A(t \mid \nu) \equiv \int_{-t}^{t} f_S(x \mid \nu) \, dx$$

$$= \frac{1}{\sqrt{\nu}\, B(\frac{\nu}{2}, \frac{1}{2})} \int_{-t}^{t} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx \qquad 71]$$

Table 5: Comparison of Bayesian and Conventional Difference of Means Test: t-statistic for m=1 and various values for n and V.

t-statistic

n \ V	0.25		1.0		4.0		16.0	
	Bayes.	Conv.	Bayes.	Conv.	Bayes.	Conv.	Bayes.	Conv.
10	4.61	6.32	2.45	3.16	.421	1.58	0	.790
20	7.44	8.94	4.01	4.47	1.59	2.24	0	1.118
30	9.62	10.95	5.08	5.48	2.24	2.74	.121	1.37
40	11.43	12.65	5.97	6.32	2.76	3.16	.641	1.58
50	13.02	14.14	6.75	7.07	3.16	3.54	1.02	1.77
60	14.45	15.49	7.44	7.75	3.55	3.87	1.28	1.94
70	15.75	16.73	8.09	8.37	3.88	4.18	1.51	2.09
80	16.96	17.89	8.68	8.94	4.18	4.47	1.70	2.24
90	18.09	18.97	9.24	9.49	4.47	4.74	1.88	2.37
100	19.16	20.00	9.76	10.0	4.74	5.00	2.04	2.50
				One tail Probability:				
10	.00037	.00006	.01448	.00575	.33735	.07415	.5	.22476
20	.00000	.00000	.00029	.00013	.05993	.01877	.5	.13874
30	.00000	.00000	.00000	.00000	.01519	.00521	.45179	.09070
40	.00000	.00000	.00000	.00000	.00402	.00151	.26108	.06096
50	.00000	.00000	.00000	.00000	.00123	.00045	.15549	.04166
60	.00000	.00000	.00000	.00000	.00034	.00013	.10062	.02880
70	.00000	.00000	.00000	.00000	.00010	.00004	.06724	.02007
80	.00000	.00000	.00000	.00000	.00003	.00001	.04537	.01408
90	.00000	.00000	.00000	.00000	.00001	.00000	.03128	.00993
100	.00000	.00000	.00000	.00000	.00000	.00000	.02170	.00703

Table 6: Comparison of Bayesian and Conventional Difference of Means Test: One-Tail p-levels for m=1 and various values for n and V.

Part I: Bayesian Difference of Means Test

	WTS-LEX WTS	WTS-LEX RANKS	WTS-LEX EGO	WTS-GDP EGO	LEX WTS-LEX RANKS	LEX WTS-LEX EGO	LEX WTS-GDP EGO	LEX RANKS-LEX EGO	LEX RANKS-GDP EGO	GDP EGO-LEX EGO
Prior n	∞	9	38	11	11	∞	1005	8	14	91
ndf	—	29	58	31	31	—	1025	28	34	111
$(V)_{tr}^*$	.0103	.1073	.0292	.0279	.0868	.0495	.0494	.0330	.0285	.0060
$(m)_{tr}^*$	.0301	.2027	.0685	.0962	.1656	.0430	.0706	-.1161	-.0879	-.0275
Posterior m	—	.1795	.0311	.0803	.1378	—	.0018	-.1068	-.0671	-.0066
t	—	2.296	1.067	2.069	2.025	—	.2031	2.416	1.796	.6844
P one-tail	.5	.0145	.1453	.0235	.0258	.5	.4196	.0112	.0407	.2476

Part II: Conventional Difference of Means Test

	WTS-LEX WTS	WTS-LEX RANKS	WTS-LEX EGO	WTS-GDP EGO	LEX WTS-LEX RANKS	LEX WTS-LEX EGO	LEX WTS-GDP EGO	LEX RANKS-LEX EGO	LEX RANKS-GDP EGO	GDP EGO-LEX EGO
ndf	20	—	—	—	—	—	—	—	—	20
t	1.361	2.835	1.837	2.637	2.575	.8858	1.455	2.930	2.388	1.623
P one-tail	.0943	.0051	.0406	.0079	.0090	.1931	.0806	.0041	.0135	.0601

* The subscript "tr" denotes the value of the sample variance and mean after each observation x_i was transformed to the infinite interval according to the expression $(x_i)_{tr} = \tan\left(\frac{\Pi x_i}{4}\right)$

Examples

Two sets of results are presented as illustration of the Bayesian, one sample "t-test" developed above. The first set of results (Table 5) merely compares the Bayesian and conventional t-test for a sample mean of 1.0 and a sample standard deviation ranging from 0.5 to 4.0. Note that the Bayesian p-levels are always conservative relative to the conventional values. Indeed, for large variance and small n the Bayesian p-level is about 1/3 - 1/4 the value of the conventional figure.

However, the real power of the Bayesian approach is shown in Table 6, where the posterior mean gives a conservative estimate of the effect under study. The data for this example come from a study of two varieties of models used to describe individual choice behavior (Lavin, 1969). The first type, called the weights model, is denoted by WTS in Table 6. In this model, the decision maker is supposed to weight all the attributes, score each alternative on each attribute, and then base his choice on the sum of weighted attribute scores.

In the second type of model the decision maker is required to do much less computation. He merely ranks the attributes and then ranks alternatives on the most important attribute. Lower order attributes are used only to break ties. Table 6 uses the symbol LEX to characterize the lexicographic structure of this model. The second element of a LEX model's name refers to the source of attribute ranks: WTS and RANKS denote sets of ranks which are based on direct queries about attributes weights and ranks, respectively; EGO denotes an indirect measure of importance using Sherif's (1965) notion of ego-involvement.

Results are also reported for a variety of the lexicographic model which uses aspiration level data to dichotomize attributes into selection criteria and constraints. The symbol GDP stands for Generalized Decision Processor, the name given to the non-routine decision program in which this type of lexicographic choice routine appears. (Soelberg, 1967)

The comparisons made in Table 6 are based on the differences in rank order correlations between predicted and actual alternative ranks. Because the coefficient of correlation ranges between -1 and +1, the correlation differences must first be transformed to an infinite interval in order to be consistent with the assumed normality of the underlying distribution.

An examination of Table 6 reveals some interesting differences between the two tests. When there is an unmistakable difference, as in the WTS-LEX RANKS comparison, the Bayesian test produces a p-level which is pleasingly small, though nonetheless three times larger than the conventional value. But more important, the Bayesian analysis states that 0.18 is a conservative estimate of the difference in model accuracy, a difference of considerable "practical significance."

In a similar vein, we note that the WTS model is significantly more accurate than the LEX EGO model at the 0.05 level of the conventional test; however, this difference is much less significant when viewed with Bayesian eyes. Indeed, a posterior estimate of the mean difference in precision is only 0.03 which is hardly impressive by any standards.

In the next section we develop the last test in the paper by extending the results of section 5 to the case of two independent samples.

6. Two Sample Difference of Means Test

The usual difference of means problem requires one to infer the mean difference between two populations when one knows neither the mean nor the variance in either population. Suppose one has two independent random samples $(x_{11}, x_{12}, \ldots x_{1n})$ and $(x_{21}, x_{22}, \ldots x_{2n_2})$ of size n_1 and n_2, respectively, and suppose that the parameter of interest, x, is normally distributed with means μ_1, μ_2 and precisions h_1, h_2 in the two populations.

The conditional sampling distribution is

$$P_r\{\underline{x}_1, \underline{x}_2 \mid n_1, n_2, \mu_1, \mu_2, h_1, h_2\} =$$

$$= \left(\frac{h_1}{2\pi}\right)^{n_1/2} \left(\frac{h_2}{2\pi}\right)^{n_2/2} \exp\left\{-\frac{h_1}{2}\sum_{i=1}^{n_1}(x_{1i}-\mu_1)^2\right\} \exp\left\{-\frac{h_2}{2}\sum_{j=1}^{n_2}(x_{2j}-\mu_2)^2\right\} \tag{72}$$

and the max likelihood estimators are

$$\hat{\mu}_i = m_i \equiv \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

$$\hat{h}_i = 1/s_i^2 \equiv \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij}-m_i)^2 \quad , \quad (i=1,2) \tag{73}$$

Thus, using the one-sample analysis as a guide, we select a multivariate prior which is a product of two normals and two gammas:

$$f'(\mu_i, h_i) = f_{N1}(\mu_1 \mid m_1', n_1' h_1)\ f_{\gamma 2}(h_1 \mid \nu_1', V_1')\ f_{N1}(\mu_2 \mid m_2', n_2' h_2)\ f_{\gamma 2}(h_2 \mid \nu_2', V_2'),$$

$$(\nu_1', \nu_2' > 1) \tag{74}$$

Clearly, the posterior will be of the same form because we have chosen a prior which is conjugate to the conditional sampling distribution. One may immediately write the posterior parameters by subscripting the one sample results of 59] .

$$n_i'' \equiv n_i + n_i'$$
$$\nu_i'' \equiv \nu_i + \nu_i' + 1$$
$$\nu_i \equiv n_i - 1\ ,\quad \nu_i' = n_i' - 1$$
$$m_i'' \equiv (m_i n_i + m_i' n_i')/n_i''$$
$$V_i'' \equiv \left[\nu_i V_i + \nu_i' V_i' + \frac{n_i n_i'}{n_i''}(m_i - m_i')^2\right]/\nu_i'' \qquad ,\ (i=1,2) \qquad 75]$$

Similarly, the unconditional sampling distribution is merely the product of two one-sample expressions:

$$\mathbb{P}(m_i, V_i \mid m_i', V_i', n_i', \nu_i')$$
$$= (2\pi)^{-\frac{n_1+n_2}{2}} \left(\frac{n_1'}{n_1''}\right)^{1/2} \left(\frac{n_2'}{n_2''}\right)^{1/2} \frac{\Gamma\left(\frac{n_1''-1}{2}\right)}{\Gamma\left(\frac{n_1'-1}{2}\right)} \frac{\Gamma\left(\frac{n_2''-1}{2}\right)}{\Gamma\left(\frac{n_2'-1}{2}\right)} \frac{\left(\frac{\nu_1' V_1'}{2}\right)^{\nu_1'/2}}{\left(\frac{\nu_1'' V_1''}{2}\right)^{\nu_1''/2}} \frac{\left(\frac{\nu_2' V_2'}{2}\right)^{\nu_2'/2}}{\left(\frac{\nu_2'' V_2''}{2}\right)^{\nu_2''/2}} \qquad 76]$$

To determine prior parameters, we translate the "no difference" idea of the null hypothesis into the condition

$$m_1' = m_2' \equiv m' \qquad 77]$$

which constrains the internal extrema of 72] . Upon setting the partial derivatives of $\ln \mathbb{P}$ to zero and applying 77] one obtains the

following set of simultaneous equations:

$$0 = \frac{\nu_i'}{2}\frac{1}{V_i'} - \frac{\nu_i''}{2}\frac{1}{V_i''}\frac{\partial V_i''}{\partial V_i'}$$

$$0 = \frac{1}{n_i'} - \frac{1}{n_i''} + \left[\psi\left(\frac{n_i+n_i'-1}{2}\right) - \psi\left(\frac{n_i'-1}{2}\right)\right] + \ln\left(\frac{n_i'-1}{n_i+n_i'-1}\right) + \ln\left(\frac{V_i'}{V_i''}\right)$$
$$- \frac{n_i+n_i'-1}{V_i''}\frac{\partial V_i''}{\partial n_i'}$$

$$0 = \frac{\nu_1''}{2}\frac{1}{V_1''}\frac{\partial V_1}{\partial m'} + \frac{\nu_2''}{2}\frac{1}{V_2''}\frac{\partial V_2''}{\partial m'} \qquad (i = 1, 2) \qquad 78\text{ a,b,c]}$$

With the aid of 75] and 77], these equations can be reduced to the set

$$V_i' = \frac{\nu_i}{n_i}V_i + \frac{n_i}{n_i''}(m_i - m')^2$$

$$m' = \frac{(n_1 n_1' n_2'' V_2')m_1 + (n_2 n_2' n_1'' V_1')m_2}{(n_1 n_1' n_2'' V_2') + (n_2 n_2' n_1'' V_1')}$$

$$0 = \frac{n_i}{n_i' n_i''} + \left[\psi\left(\frac{n_i+n_i'-1}{2}\right) - \psi\left(\frac{n_i'-1}{2}\right)\right] + \ln\left(\frac{n_i'-1}{n_i+n_i'-1}\right) - \left(\frac{n_i}{n_i''}\right)^2\frac{(m_i-m')^2}{V_i'}$$

79 a,b,c]

As in the one sample case, our particular choice of a prior causes the prior and posterior variance to be equal for each population.

The set 79] can be reduced to three equations by the simple artifice of substituting for V_i' in 79 b,c] :

$$0 = \frac{1}{n_1''}\frac{n_1 n_1'(m_1 - m')}{\frac{n_1-1}{n_1}V_1 + \frac{n_1'}{n_1''}(m_1-m')^2} + \frac{1}{n_2''}\frac{n_2 n_2'(m_2-m')}{\frac{n_2-1}{n_2}V_2 + \frac{n_2'}{n_2''}(m_2-m')^2}$$

$$0 = \frac{n_i}{n_i' n_i''} + \left[\psi\left(\frac{n_i+n_i'-1}{2}\right) - \psi\left(\frac{n_i'-1}{2}\right)\right] + \ln\left(\frac{n_i'-1}{n_i+n_i'-1}\right) - \left(\frac{n_i}{n_i''}\right)^2\frac{(m_i-m')^2}{\frac{n_i-1}{n_i}V_i + \frac{n_i'}{n_i''}(m_i-m')^2}$$

80 a,b]

This is quite a simplification. But we can boil 80a,b] down a bit more by making the following definitions:

$$a_i \equiv \frac{n_i}{n_i' n_i''} + \left[\psi\left(\frac{n_i+n_i'-1}{2}\right) - \psi\left(\frac{n_i'-1}{2}\right)\right] + \ln\left(\frac{n_i'-1}{n_i+n_i'-1}\right)$$

$$b_i \equiv \left(\frac{n_i}{n_i''}\right)^2$$

$$c_i \equiv \frac{n_i-1}{n_i} V_i$$

$$d_i \equiv \frac{n_i}{n_i'} \qquad (i = 1,2) \qquad 81]$$

Equations 80b] can now be rewritten as

$$0 = a_i - \frac{b_i (m_i - m')^2}{c_i + d_i (m_i - m')^2}$$

which yield the expressions

$$(m_i - m') = \left[a_i c_i / (b_i - a_i d_i)\right]^{1/2} \equiv \Lambda_i^{1/2} \qquad 82]$$

Upon substituting 82] into 80a] and eliminating m' from 82], we obtain two simultaneous equations in n_1' and n_2'.

$$0 = \frac{n_1 d_1 \Lambda_1^{1/2}}{\frac{n_1-1}{n_1} V_1 + d_1 \Lambda_1} + \frac{n_2 d_2 \Lambda_2^{1/2}}{\frac{n_2-1}{n_2} V_2 + d_2 \Lambda_2}$$

$$m_2 - m_1 = \Lambda_2^{1/2} - \Lambda_1^{1/2} \qquad 83a,b]$$

Equations 83] exhibit a most convenient structure, namely,

$$u_1(n_1') + u_2(n_2') = 0$$

$$v_1(n_1') - v_2(n_2') = \Delta m \equiv m_1 - m_2 \qquad \text{84 a,b]}$$

We see no way to simplify these equations further. Nonetheless, it appears that a numerical solution would be quite straightforward. One merely plots two curves in the n_1'-n_2' plane -- one for which $u_1 + u_2 = 0$ and one for which $v_1 + v_2 = \Delta m$. Their intersection defines the values of n_1' and n_2' for an internal maximum (c.f. Figure 3)

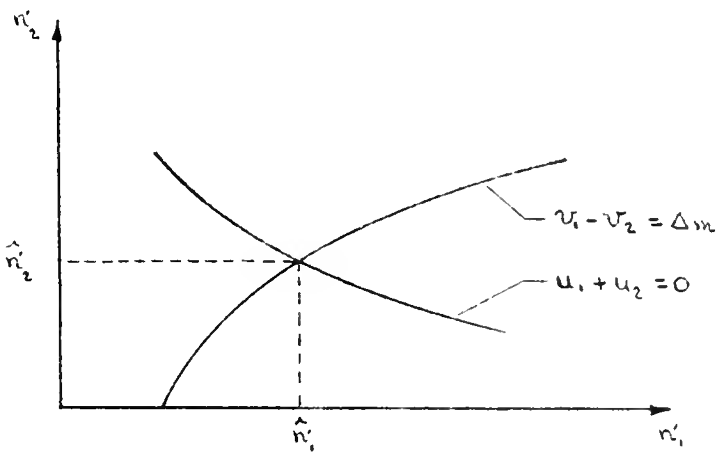


Figure 3: Sketch of a Numerical Method to Solve for Prior Parameters when $\mathbb{P}$ has an Internal Maximum

Now that the prior parameters have been determined, all that remain to be computed are the desired posterior properties. Although the variances in the two populations are interesting, it seems to us that the researcher is invariably keen to know certain statistics of the difference between the means,

$$\tilde{z} \equiv \mu_2 - \mu_1 \tag{85}$$

In particular, we expect him to be interested in the posterior mean and variance of $\tilde{z}$ -- $\bar{z}''$ and $\check{z}''$ respectively - and $P_r\{\tilde{z} \geq \alpha\}$.

Of course, the calculation of each of these quantities begins with the joint distribution on μ_1, μ_2, h_1, h_2. Examination of 74] indicates that the joint is factorable into two parts, each of which resembles the joint in the one-sample case. Thus, the marginal on μ_1, μ_2 is a product of two non-central t-distributions:

$$f''(\mu_1, \mu_2 | \underline{x}_1, \underline{x}_2) = f_{S_1}(\mu_1 | m_1'', \frac{n_1''}{V_1''}, \nu_1'') \; f_{S_2}(\mu_2 | m_2'', \frac{n_2''}{V_2''}, \nu_2'') \tag{86}$$

It is a straightforward matter to compute $\bar{z}''$ and $\check{z}''$, so we shall merely state the results:

$$\bar{z}'' = m_2'' - m_1''$$

$$\check{z}'' = \frac{V_2''}{n_2''}\frac{\nu_2''}{\nu_2''-2} + \frac{V_1''}{n_1''}\frac{\nu_1''}{\nu_1''-2} \tag{87 a,b}$$

The calculation of the tail probability is much more challenging.

$$P_r\{\tilde{z} \geq \alpha\} = 1 - P_r\{\tilde{z} < \alpha\}$$

$$= 1 - \int_{-\infty}^{\infty} dv\, f_{S_1}(v) \int_{-\infty}^{v+\alpha} du\, f_{S_2}(u) \tag{88}$$

After the transformation

$$(u-m_2)(n_2/V_2)^{1/2} = \nu_2 \tan\theta$$

equation 88] becomes

$$P_r\{z>\alpha\} = 1 - \frac{1}{B(\frac{\nu_2}{2},\frac{1}{2})} \int_{-\infty}^{\infty} dv\, f_{s_1}(v) \int_{-\pi/2}^{\beta} (\cos\theta)^{\nu_2-1}\, d\theta \qquad 89]$$

where

$$\beta \equiv \tan^{-1}\left\{(v+\alpha-m_2)(n_2/\nu_2 V_2)^{1/2}\right\}$$

and all primes have been omitted to avoid clutter. After applying formula 331.7a) in Grobner and Hofreiter and simplifying, we obtain the final result:

$$P_r\{z>\alpha\} = 1 - \mathbb{C}\int_{-\infty}^{\infty} dv\left[\frac{n_1}{V_1}(v-m_1)^2+\nu_1\right]^{-\frac{\nu_1+1}{2}}\left\{(\sin\beta)\sum_{k=1}^{r} a_k(\cos\beta)^{\nu_2-2k} + b\left(\beta-\frac{\pi}{2}\right)\right\} \qquad 90]$$

where

$$\cos\beta \equiv \left\{1+\frac{n_2}{\nu_2 V_2}(v+\alpha-m_2)^2\right\}^{-1/2}$$

$$b \equiv (1-s;2;r)/(2-s;2;r)$$

$$\nu_2 - 1 = 2r-s, \quad s = 1 \text{ or } 0$$

$$a_k \equiv (\nu_2-2;-2;k-1)/(\nu_2-1;-2;k)$$

$$\mathbb{C} \equiv \left(\frac{n_1}{V_1}\right)^{1/2}\frac{\nu_1}{B(\frac{\nu_1}{2},\frac{1}{2})\,B(\frac{\nu_2}{2},\frac{1}{2})}$$

and ν_1, ν_2, n_1, n_2 are restricted to integral values. Evidently, 90] must be evaluated numerically.

This completes the development of the last test in the program announced above. We apologize for failing to include a numerical example; we simply lacked the patience to compute the prior parameters.

In the following section we review the justification for the approach we have taken and comment on the philosophical implications of using the data to define a prior.

7. Discussion

The fundamental tenet of this analysis is that hypothesis testing is essentially an attempt to estimate the strength of an effect. Because an effect is usually measured in terms of a difference in proportions or means that can never be known precisely, any estimate must be based on a diffuse hypothesis about the parameters of interest. Such a diffuse hypothesis is merely a probability density function, a quantification of our belief about where the true value of the parameter lies.

Clearly, the conditional law of probability provides a natural schema to make inferences about parameters if one can unambiguously define an acceptable prior hypothesis. The problem confronted in this paper is development and application of a rule for choosing a prior hypothesis which is both unambiguous and conservative. The rule can be stated very simply: choose the prior hypothesis so that 1) its expectation is the conventional null value and 2) it has maximum probability of producing the observed data. For mathematical convenience, the form of the prior is tailored to combine simply with the conditional sampling distribution. If a conjugate prior seems too constraining, one can always select a more satisfying function; however, the above rule for defining the exact form of the prior still applies.

At a more abstract level, the choice of a prior has been posed as a problem in constrained maximization. Although we have imposed only one constraint, future research may suggest others. For example, one could formalize the fortuitous result of the difference of means tests by requiring that the prior and posterior variance be the same. Of course, the greater the number of constraints, the lower the probability that the chosen prior produced the data. So we prefer to keep the constraints to a minimum.

The idea of using the data to define a prior is novel if not a bit shocking. Indeed, some may assert that an hypothesis is no longer a <u>prior</u> hypothesis if its parameters are based on the data about which we propose to make a posteriori inferences. But note that we do specify the rule in advance of observation much as a careful researcher spells out his operational definitions before subjecting the data to analysis. Moreover, the rule is constructed to produce a null (prior) which is as hostile to the research hypothesis as possible, <u>in light of the observed data</u>. We shall let Clutton-Brock (1965, p. 27) have the closing word:

> In my view, we are entitled to use any evidence, prior or posterior to form an estimate of the prior frequency distribution. I prefer to retain the word "prior" in my description of the estimate, because I feel that the name of what is estimated should not be altered by the method of estimation.

References

Clutton-Brock, M. "Using the Observations to Estimate the Prior Distribution," Journal of the Royal Statistical Society, Series B, 27:17-27 (1965).

Day, R.H. and Power, R.P., "Apparent Reversal (Oscillation) of Rotary Motion in Depth: An Investigation and a General Theory" Psychological Review 72:117-127 (1965).

Edwards, W., Lindman, H., and Savage, L.J., "Bayesian Statistical Inference for Psychological Research," Psychological Review, 70:193-242 (1963).

Good, I.J., The Estimation of Probability: An Essay on Modern Bayesian Methods, Research Monograph no. 30, The MIT Press, Cambridge, Mass. (1965).

Gradshtein, I.S., and Ryzhik, I.M., Table of Integrals, Series, and Products, 4th edition (English translation), Academic Press, New York, (1965).

Gröbner, W. and Hofreiter, N., Integraltafel: Erster Teil, Unbestimmte Integrale; Zweiter Teil, Bestimmte Integrale, Springer-Verlag, Wien (1961)

Hershberger, W.A., "Comment on 'Apparent Reversal (Oscillation) of Rotary Motion in Depth'", Psychological Review 74:235-238 (1967).

Jaynes, E.T., "Prior Probabilities," IEEE Transactions on System Science and Cybernetics, Vol. SSC-4, no. 3, pp. 227-241 (Sept. 1968).

Lavin, M.L., "A Comparison of Descriptive Choice Models," unpublished Ph.D. dissertation, Sloan School of Management, MIT, Cambridge, Mass. (1969).

Pitz, G.F., "An Example of Bayesian Hypothesis Testing: The Perception of Rotary Motion in Depth," Psychological Bulletin, 70:252-255 (1968).

Raiffa, H. and Schlaiffer, R., Applied Statistical Decision Theory, Division of Research, Graduate School of Business Administration, Harvard University, Boston (1961).

Soelberg, P.O., "A Study of Decision Making: Job Choice," unpublished Ph.D. dissertation, Graduate School of Industrial Administration, Carnegie-Mellon University (1967).

Wilensky, H.L., "The Labor Vote: A Local Union's Impact on the Political Conduct of the Members," Social Forces, 35:111-120 (1956).